The crystal structure of diphtheria toxin

Seunghyon Choe*, Melanie J. Bennett*, Gary Fujii, Paul M. G. Curmi, Katherine A. Kantardjieff, R. John Collier† & David Eisenberg*

* Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90024, USA
† Department of Microbiology and Molecular Genetics, Harvard Medical School, Shipley Institute of Medicine, Boston, Massachusetts 02115, USA

The crystal structure of the diphtheria toxin dimer at 2.5 Å resolution reveals a Y-shaped molecule of three domains. The catalytic domain, called fragment A, is of the $\alpha+\beta$ type. Fragment B actually consists of two domains. The transmembrane domain consists of nine α-helices, two pairs of which are unusually apolar and may participate in pH-triggered membrane insertion and translocation. The receptor-binding domain is a flattened β-barrel with a jelly-roll-like topology. Three distinct functions of the toxin, each carried out by a separate structural domain, can be useful in designing chimaeric proteins, such as immunotoxins, in which the receptor-binding domain is substituted with antibodies to target other cell types.

DIPHTHERIA was a major cause of death among children until mass immunization against diphtheria toxin (DT) reached the general public in the late 1920s. DT is secreted as a single polypeptide chain of 535 residues from *Corynebacterium diphtheriae* that has been lysogenized by a bacteriophage carrying the DT gene[1]. Mild trypsinization and reduction of DT *in vitro* generates two fragments, fragment A (amino terminal, about 21K) and fragment B (carboxy terminal, about 37K), as a result of cleavage at residue 190, 192 or 193 (refs 2, 3). A similar proteolytic cleavage ('nicking') occurs *in vivo* before or soon after the toxin binds to a sensitive cell[4]. Fragment B of the toxin binds the protein to receptors on the cell surface and promotes transfer of fragment A to the cytoplasm. Fragment A in the cytoplasm catalyses the transfer of the ADP-ribosyl group of NAD^+ to elongation factor 2 (EF-2). This inactivates EF-2, stopping protein synthesis and killing the target cell. Introduction of a single molecule of fragment A into the cytoplasm can kill a cell[5]. Studies are underway to create therapeutically active chimaeric proteins, with a part of DT linked to an antibody or other ligand that can recognize and deliver the toxin's lethal activity to a specific target cell[6–9].

Beyond its interest as a toxic enzyme, DT has been studied to understand the mechanism by which it translocates its enzymatically active fragment A into cells. After binding to a receptor, the toxin is endocytosed and delivered to endosomes, where conditions are acidic. At a threshold pH of ~5.0 the toxin undergoes a conformational change, which promotes insertion and formation of an ion-selective channel in the membrane, and fragment A is translocated and released into the cytoplasm[10–12]. The relationship between pore formation and transfer of fragment A across the endosomal membrane remains uncertain. Hydrophobic segments in the central part of fragment B have been implicated in the channel formation[13,14]. The membrane translocation machinery of DT, therefore, can be considered as a simple translocation system consisting of a single protein, which may illuminate the protein secretory apparatus of cells.

Structure determination

The structure is based on analyses of form 1, form 3 and form 4 crystals. Form 1 crystals of DT complexed with adenylyl-3',5'-uridine monophosphate (ApUp) belong to triclinic space group $P1$ with unit cell dimensions of $a=70.4$ Å, $b=70.6$ Å, $c=65.4$ Å, $\alpha=94.9°$, $\beta=91.0°$ and $\gamma=99.6°$, with two chains per asymmetric unit. This dimeric asymmetric unit is consistent with the discovery, after our initial report of crystallization[15], that the crystals were of a dimeric form of DT sometimes found in crude or purified preparations of the protein. Dimeric DT itself is not toxic, presumably because it does not bind to receptors, but it slowly dissociates to fully toxic monomers[16]. The dimer may represent a conformationally altered form of the biologically active monomeric toxin. Irreproducible crystallization conditions for obtaining form 1 crystals hampered crystallographic studies of structure determination until three new crystal forms were obtained[17]. Form 3 and form 4 belong to monoclinic space group $C2$, with unit cell dimensions for form 3 of $a=107.3$ Å, $b=91.7$ Å, $c=66.3$ Å and $\beta=94.7°$, and for form 4 of $a=108.3$ Å, $b=92.3$ Å, $c=66.1$ Å and $\beta=90.4°$. In both of these forms there is one DT chain per asymmetric unit and pairs of DT chains are related by a 2-fold rotation axis.

The initial model was based on the structure determination of form 4 crystals at 3.0 Å resolution, using the multiple isomorphous replacement (MIR) method followed by solvent flattening[18]. With the initial model, the structures of form 1 and form 3 were readily solved by molecular replacement[19,20]. Single isomorphous replacement (SIR) phases were also obtained for form 3. Native data were then collected to 2.5 Å resolution, and the model was rebuilt into 2.5 Å maps with form 3 (SIR) and form 4 (MIR) after the phases had been extended and modified by the method of Zhang and Main[21]. This was followed by real-space density averaging between two forms. Sequence fitting was difficult in the ~120 C-terminal residues (part of the receptor-binding or R domain) where the most ambiguous regions were near residues 408 and 510. Some of the useful markers in the density maps were Trp 50, Trp 153, Trp 281, Trp 398, a 5-residue segment of Met 178, Tyr 179, Glu 180, Tyr 181, Met 182, a 4-residue cluster of Phe 355, Tyr 358, His 372, Tyr 375, a cluster of Tyr 514, Phe 530, Phe 531 with big side chains near the C terminus (Fig. 1*c*), and two disulphide bonds between Cys 186 and Cys 201 and Cys 461 and Cys 471. An initial improper fitting in the R domain was detected by profile window plots[22] and then corrected. Iterative cycles of refinement were done independently at 2.5 Å for each dataset. The atomic model for each form is essentially identical except for crystal packing. Details of phase modification and refinement will be described elsewhere. Assessment of the accuracy of the model rests on the fit of the model to the MIR and density-modified maps, crystallographic R factors, real-space R factors[23], the free R value[24], which is only 4% higher than the crystallographic R factor, and profile window plots[22]. At the present stage of refinement, the agreement of the atomic models to crystallographic data is characterized by R factors of 21.1, 21.6 and

Present address: Vestar, Inc., San Dimas, California 91773, USA (G.F.); School of Physics, University of New South Wales, Kensington, Australia (P.M.G.C.); Department of Chemistry and Biochemistry, California State University, Fullerton, California 92634, USA (K.A.K.).

TABLE 1 Statistics for X-ray data collection, phase determination and refinement

Native data		Overall (R^*_{scale})	10–4.0	4.0–3.5	3.5–3.0	3.0–2.5 Å
Form 4	Total	36,758 (11.9)	26,897	4,977	4,884	
	Unique (% complete)	10,875 (83)	5,190 (99)	2,414 (92)	3,271 (63)	
From 4 (new)	Total	35,897 (6.1)				
	Unique (% complete)	18,665 (84)	5,195 (99)	2,673 (98)	4,268 (82)	6,529 (72)
Form 3	Total	61,009 (7.6)	21,984	15,368	10,573	13,084
	Unique (% complete)	19,912 (90)	5,231 (100)	2,682 (98)	4,603 (88)	7,396 (82)
Form 1	Total	66,464 (7.5)	22,245	21,118	15,277	7,824
	Unique (% complete)	25,854 (68)	6,523 (96)	7,665 (92)	7,102 (76)	4,574 (35)
Derivatives		Overall	10–4.6	4.6–3.6	3.6–3.0	3.0–2.8
Form 4						
KOS	Unique (R^*_{scale})	11,765 (9.16)				
	$R^\dagger_c$(*fh/e*)	0.66 (1.23)	0.66 (1.29)	0.62 (1.18)	0.75 (1.28)	0.80 (1.06)
CNP	Unique (R^*_{scale})	12,255 (12.0)				
	$R^\dagger_c$(*fh/e*)	0.70 (1.06)	0.68 (1.33)	0.72 (0.93)	0.72 (0.97)	0.66 (1.11)
KNP	Unique (R^*_{scale})	8,164 (8.32)				
	$R^\dagger_c$(*fh/e*)	0.71 (1.00)	0.72 (0.87)	0.67 (1.18)	0.75 (1.33)	
CAP	Unique (R^*_{scale})	7,552 (15.3)				
	$R^\dagger_c$(*fh/e*)	0.71 (1.28)	0.70 (1.54)	0.72 (1.12)	0.88 (1.20)	
KAP	Unique (R^*_{scale})	10,152 (12.26)				
	$R^\dagger_c$(*fh/e*)	0.81 (1.26)	0.81 (1.43)	0.71 (1.19)	0.75 (0.89)	
GCL	Unique (R^*_{scale})	6,595 (11.90)				
	$R^\dagger_c$(*fh/e*)	0.70 (1.10)	0.69 (1.13)	0.66 (1.09)	0.50 (1.05)	
Form 3						
KOS	Unique (R^*_{scale})	11,435 (13.57)				
	$R^\dagger_c$(*fh/e*)	0.54 (1.10)	0.56 (1.32)	0.60 (0.74)		
Refinement		Form 1	Form 3	Form 4		
	R factor§ (6–2.5 Å)	0.211	0.216	0.219		
	r.m.s. bond (Å)	0.021	0.021	0.021		
	r.m.s. angle	4.54	4.40	4.48		
	r.m.s. dihedral	26.4	25.9	26.1		

Crystal forms 1, 3 and 4 were used for the current study[17]. Diffraction data were collected on a Rigaku AFC-6 diffractometer operating at 8.5 kW, equipped with a two-panel area detector of Xuong-Hamlin design (San Diego Multiwire Systems). Images were recorded as 0.1° oscillation frames, integrated and merged into batches of 50 frames (5°). Integrated intensities were scaled and merged by the FOURIER scaling method[42]. Form 4 native and derivative data were later collected to 2.5 Å with a RAXIS imaging plate system. Heavy atom derivatives: KOS, K_2OsO_4, soaked for 3 days at saturated concentration in artificial mother liquor (12% PEG 8000, 0.43 M NaCl, 43 mM Tris-HCl, pH 7.8); CNP, 4-chloro-2-nitro-mercury phenol, soaked for 5 days at saturated concentration in artificial mother liquor; KNP, 1 to 1 mixture of KOS and CNP; CAP, trans-dichlorodiamine Platinum (II), soaked for 3 days at 2 mg ml^{-1} in artificial mother liquor; KAP, 1 to 1 mixture of KOS and CAP; GCL, $HgCl_2$, soaked for 3 days at 2 mg ml^{-1} in artificial mother liquor. Heavy atom parameters were refined and MIR phases calculated using the program HEAVY[43]. We initially obtained the Os derivative for form 3 crystals. From electron density maps based on the SIR phases after solvent flattening at 3.5 Å resolution, the shape of the molecule was interpreted to have three domains. But secondary structures were not easily interpretable and the course of the polypeptide chain was difficult to determine. A search for additional heavy atom derivatives was hampered by the lack of good quality crystals of form 3. We therefore shifted our efforts to form 4 crystals. MIR phases for form 4 were obtained from six heavy atom derivatives using isomorphous differences and anomalous differences. The Os and Pt derivatives were solved by isomorphous difference Patterson functions, and the Hg derivative by a difference Fourier synthesis. Os derivatives of form 4 and form 3 have the same single site binding. Solvent flattening: initial electron density maps of form 4 were calculated at 3.0 Å resolution, with phases modified using an iterative solvent flattening procedure[18] including phases extended to 3.0 Å from 3.2 Å by the Wang phase extension algorithm[18]. A solvent volume of 45% was used to ensure that all protein density was included in the protein mask, somewhat smaller than the 57% estimated from the M_r. From these maps, all secondary structures were identified and an initial model was built using a polyalanine chain. Model building was expedited with the program FRODO[44] and the fragment-fitting routines of the program O[23]. Starting with α-carbon coordinates that were manually built, main chain atoms were added using a database of 34 well refined protein structures. Then side chains were added using a rotamer database[45]. Refinement: this initial model was adjusted by visual inspection of density maps before it was refined by the simulated annealing protocol of the program XPLOR[46]. The relative orientations of DT in forms 1, 3 and 4 were determined by a Patterson-space rotation and translation search of the refined form 4 model against form 1 and form 3 data. Two top solutions (9σ) for form 1 data correspond to two DT chains related by noncrystallographic symmetry in the asymmetric unit. The transformation from form 4 to form 1 is essentially a change of coordinate system from $C2$ to $P1$, where the crystallographic rotation axis of $C2$ becomes a noncrystallographic rotation symmetry axis of $P1$ that is nearly parallel to the (110) axis of $P1$. One top solution (7σ) for form 3 corresponds to a rotation of less than 0.5° in all three directions. The transformation from form 4 to form 3 is essentially a 5 Å translation along the *a* axis. This result is consistent with the observation that the average absolute difference of the amplitudes of structure factors of 0kl reflections between form 3 and form 4 is 15%, whereas those between h0l or hk0 reflections between form 3 and form 4 are almost random ($R = 48\%$). Also, when the model was superimposed on the solvent-flattened electron density maps of form 3 based on the SIR phases, most of the secondary structures were recognized with the model as a guide. Real-space averaging of densities between form 4 and form 3 with MIR and SIR phases at 3.0 Å improved the density maps at this stage. Subsequently, experimental phases were extended to 2.5 Å by the algorithm based on solvent flattening, histogram matching, and Sayre's equation[21] for form 3 and form 4. Form 3 maps at 2.5 Å were again skewed and averaged with form 4 maps. These were the most interpretable maps. Refinement of the atomic model was done independently for form 1, form 3 and form 4 with all observed data having F_{obs} greater than $1\sigma(F_{obs})$ between 6 and 2.5 Å.

* $R_{scale} = \sum(|I_i - I_j|)/\sum(I_{av})$ where I_i and I_j are the *i*th and *j*th measurements of the equivalent reflections[42].

† R_C is the Cullis *R* factor for centric reflections.

‡ *fh/e* is the phasing power, *fh*, the mean amplitude of heavy atom structure factors divided by *e*, the r.m.s. lack-of-closure error.

§ *R* factor $= \sum(|F_{obs} - F_c|)/\sum(F_{obs})$ where F_{obs} and F_c are the structure factors observed and calculated from the model, respectively. The *R* factors for all forms increased by about 1.9% when a single temperature factor was used for all atoms.

21.9%, respectively, for form 1, form 3 and form 4 for all observed data having F_{obs} greater than $1\sigma(F_{obs})$ between 6 and 2.5 Å resolution.

The final model consists of 4,137 non-hydrogen atoms with individual isotropic temperature factors. The model also includes ApUp in the active site cleft of the catalytic (C) domain, but no solvent atoms. There are poorly defined regions in the electron density maps where main chain densities for residues 170–172, 190–195, 389–390, and 500–503 are not well defined. Residues 190–195 are part of the protease-sensitive region of the first disulphide loop, where nicking occurs; this region may be intrinsically flexible as may be the loop between the transmembrane (T) and R domains, which includes residues 389–390. Table 1 summarizes aspects of data collection, phase determination and refinement.

Structure of DT

DT consists of three abutting domains that are connected by interdomain linkers. The N-terminal C domain, middle T domain, and C-terminal R domain consist of residues 1–193, 205–378 and 386–535, respectively. Schematically, DT is Y-shaped with the base formed by the T domain, one arm of the Y formed by the C domain, and the other arm formed by the R domain. The Y is about 90 Å high, 50 Å across the top of the Y, but only 30 Å thick (Fig. 1).

Each of the three domains has a distinctive fold. The C domain is a mixed structure of eight β-strands (CB1–CB8) and seven α-helices (CH1–CH7). The eight β-strands form two β-sheets of three and five strands each. These β-sheets form a core that is surrounded by seven short helices. The overall folding of the C domain is similar to that of *Pseudomonas aeruginosa* exotoxin A (ETA), especially near the active site[25], a result that had been foreshadowed by a weak similarity in amino-acid sequences[26,27]. The folding of the active site region of *Escherichia coli* heat-labile enterotoxin also closely resembles that of ETA[28]. The T domain contains nine helices (TH1–TH9) that are folded into three helix layers, each of which is formed by two or more antiparallel helices. A similar feature was observed in the structure of the channel-forming domain of colicin A[29]. The R domain contains 10 β-strands (RB1–RB10), nine of which (RB2–RB10) build two β-sheets. These two β-sheets form a β-sandwich with a topology similar to a jelly-roll fold[30]. The three-domain organization of DT is shared by two other bacterial toxins: ETA and δ-endotoxin from *Bacillus thuringiensis*[31]. The catalytic domains of DT and ETA are the closest among all these domains in their structures and functions.

Catalytic domain. We view the C domain as being formed from two β-sheet subdomains, which subtend the active site cleft (Fig. 2). These β-sheets are oriented roughly perpendicular to each other and form the core of the domain. One subdomain consists of β-strands CB2, CB4 and CB8, surrounded by α-helices CH2, CH3, CH6 and CH7. The other subdomain consists of β-strands CB1, CB3, CB5, CB6 and CB7 surrounded by helices CH1, CH4 and CH5. The two subdomains are connected by extended loops CL1–CL4, which link the two subdomains. These four loops seem to endow the potential for flexibility or even extension to a longer and narrower shape. Conceivably the C domain can assume this partially unfolded structure during membrane translocation.

The active site cleft of the C domain, identified by the binding of the dinucleotide ApUp, is formed primarily by β-strands CB2, CB3 and CB7, the short helix CH3, and the loop CL2, and is also bounded by β-strand RB6 of the R domain. Located within the active site cleft are the following residues: Glu 148 which is believed to play a key role in catalysis[32], His 21[33] and Tyr 65[34], both of which have been implicated in NAD^+ binding, and various other residues suggested to be at or near the active site (Gly 52[32,35], Trp 50[36], Lys 39[37], Lys 474[38]). Least-squares superposition of the α-carbon coordinates of the C domains of DT and ETA yields an r.m.s. difference of 1.44 Å between 85 residues (16–33, 34–38, 49–66, 75–90, 91–96, 131–136, 147–164 of DT and 437–452, 454–458, 465–482, 493–508, 511–516, 540–545, 552–569 of ETA).

The rough position of the substrate NAD^+ in the active site can be inferred, because the dinucleotide, ApUp, binds competitively with NAD^+. The high affinity of ApUp (~0.3 nM as compared with ~8–16 μM for NAD^+[16]) may be a consequence of multiple contacts with the C domain and of salt bridges between the 3′-terminal phosphate of ApUp and the side chains of Thr 42 and Arg 458, the latter of which is a residue of the R domain. Although the structure of bound ApUp resembles that of NAD^+, there are enough differences between the covalent structures of NAD^+ and ApUp to make difficult the prediction of the conformation of NAD^+ in the cleft. But assuming that the adenine phosphate portion of NAD^+ binds in the same conformation as that of ApUp, we find that the nicotinamide ring will be positioned close to the site of the uridine ring. This

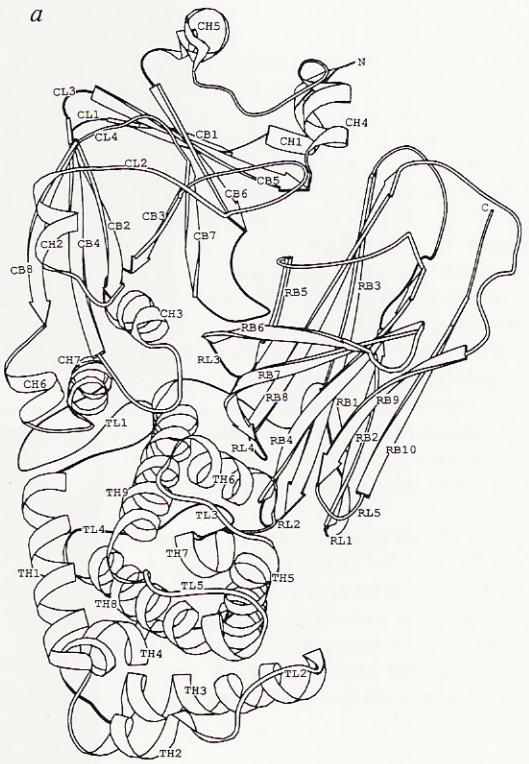
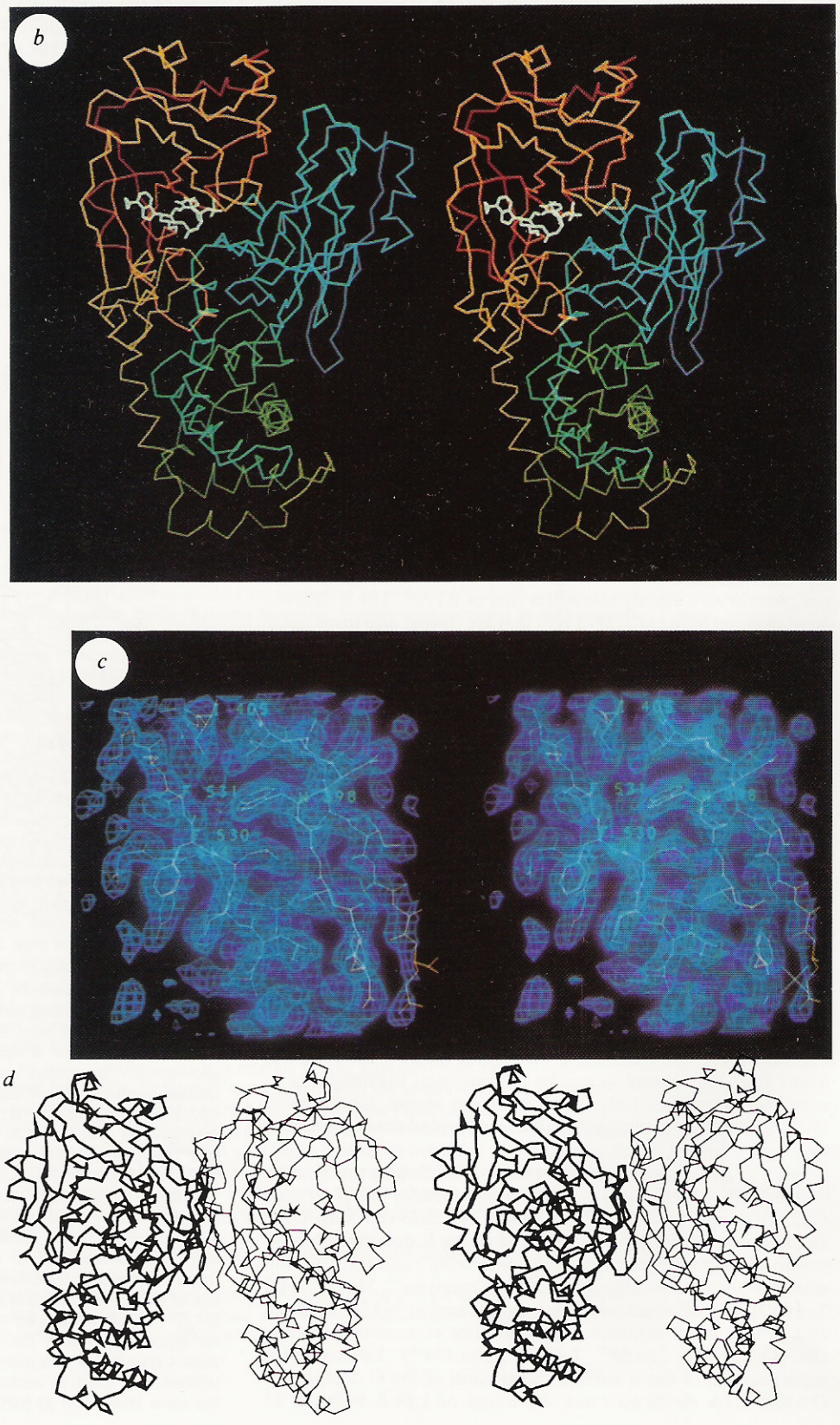


FIG. 1 *a*, Ribbon drawing of DT, labelling each secondary structural segment. The first letter denotes the domain: C for catalytic, T for transmembrane, and R for receptor-binding domains. The second letter denotes the secondary structure class: H for helix, B for β-strand, L for loop. The third symbol is the sequential number of each secondary segment from the N terminus of each domain. The ribbon program was kindly provided by J. Priestle. The residue numbers in each segment are as follows: CH1: 2–7, CB1: 11–14, CB2: 16–24, CH2: 28–34, CB3: 52–57, CH3: 58–66, CB4: 76–86, CB5: 88–96, CH4: 99–106, CH5: 120–126, CB6: 130–136, CB7: 147–152, CB8: 159–166, CH6: 168–173, CH7: 176–186; TH1: 205–221, TH2: 225–231, TH3: 238–257, TH4: 258–269, TH5: 274–288, TH6: 297–307, TH7: 310–315, TH8: 326–346, TH9: 356–378; RB1: 386–390, RB2: 393–399, RB3: 412–424, RB4: 428–438, RB5: 447–453, RB6: 455–465, RB7: 467–480, RB8: 483–495, RB9: 513–520, and RB10: 525–534. *b*, Stereo diagram of the Cα skeleton of DT from the same viewpoint as that of *a*. The colour changes gradually from red (N terminus) to blue (C terminus). An ApUp molecule (white) occupies the active site of DT. *c*, Stereo pairs of electron density maps calculated at 2.5 Å from $(2F_{obs} - F_c)$ and the refined model phases. Maps are superimposed on the corresponding region of the refined model. *d*, The DT dimer observed in the form 4 crystal. The two monomers are related by a crystallographic 2-fold rotation axis, which is vertical. The molecule at the left (in thick line) has the same orientation as that in *a*.

places the nicotinamide ring adjacent to side chains of His 21, Tyr 65 and Glu 148.

Domain junctions. One of the two intramolecular disulphide bonds of DT bridges a handle-like loop TL1 on the molecular surface (Fig. 1*a*). This 14-residue loop (187–200) connects fragment A to fragment B; it is rich in Arg and known to be easily nicked by proteases[2,3]. Once this loop is nicked, fragment A and fragment B are covalently linked only by the disulphide bond. There is evidence that nicking plays a part in the cytotoxic action of DT[4], and it is generally believed that nicked DT separates into free fragment A and fragment B when this disulphide bond is exposed to the reducing environment of the cytoplasm during



membrane translocation of the toxin. The second disulphide bond makes a nine-residue loop between residues 461 and 471 within fragment B. Residues near this loop (456, 458, 460, 472, 474) are also rich in positive charges and face the active-site cleft, probably forming the so-called phosphate-binding 'P-site' (ref. 39).

The structure suggests why whole DT is inactive in catalysing the ADP-ribosylation of EF-2 until the C domain dissociates, in the form of fragment A, from fragment B. The active site is formed at the interface between the C domain and the R domain (Fig. 1*b*). Entry to the active site is shielded by the 18-residue loop CL2 and the R domain. Thus, in whole DT, the approach of EF-2 ($M_r \sim 100$K) to the active site is blocked. The active site of whole DT remains accessible to NAD^+, however, and catalyses the NAD-glycohydrolysis (a slow side reaction that is probably physiologically insignificant). The lack of secondary structural elements within loop CL2 may allow a substantial movement of main chain atoms of the loop, permitting substrate entry to the active site.

Transmembrane domain. A central unanswered question about DT is how the low-pH environment of the endosome triggers DT insertion into the endosomal membrane and how this insertion aids the translocation of the C domain into the cytoplasm. The structure of the T domain exhibits two features that suggest how it might experience pH-triggered insertion into the membrane. The first is that the T domain is entirely α-helical, similar to the known and proposed transmembrane proteins, and that some of the helices have hydrophobic characteristics more typical of transmembrane helices than of globular proteins[40]. The nine helices are arranged more or less in three layers, each layer consisting of an antiparallel pair of helices. The two long, C-terminal helices, TH8 and TH9, are unusually apolar and constitute the central core layer. One flanking layer, made up of helices TH5–TH7, also contains hydrophobic helices, TH6 and TH7. The other layer, made up of helices TH1–TH3 is, in contrast, very hydrophilic even compared with globular proteins. The second noteworthy feature of the T domain is the acidic composition of the loops that connect pairs of these helices: both loop TL3 between helices TH5 and TH6, and loop TL5 between hydrophobic helices TH8 and TH9 contain a total of six Asp and Glu residues (Fig. 3*a*). At neutral pH, these loops are highly charged and water-soluble. But at acidic pH, these residues would be at least partially protonated, and hence more nearly neutral and membrane-soluble, especially near the surface of the membrane that has even higher concentration of protons due to the surface potential[41]. Thus the lower pH inside the endosome would tend to render these tip-shaped loops into membrane-soluble 'daggers' that would lead the two apolar helix pairs into the membrane.

Other structural characteristics of the T domain suggest that it has the capacity to insert into the membrane and can assist the translocation of the C domain. The first is that the nearly parallel packing of the three helix layers would permit spreading on the membrane surface of the first helix layer (TH1–TH3) if other layers were inserted. This insertion would require local conformational changes in loops, but no alteration of the helices themselves. Also the pronounced hydrophobic asymmetry is compatible with the proposed rearrangement: 15 of 16 Lys and Arg residues and all six His residues of the T domain are located on the opposite side from the dagger tips (Fig. 3*b*), making the whole domain a hydrophobic dipole once the Asp and Glu residues are neutralized. In short, we propose that the hairpin loop TL5 and probably TL3 cross the membrane, where the Asp and Glu residues will once again be charged in the neutral pH of the cytoplasm.

Receptor-binding domain. The R domain is formed from two β-sheets. β-Strands RB2, RB3, RB5 and RB8 form a four-stranded β-sheet that faces a five-stranded β-sheet containing β-strands RB4, RB6, RB7, RB9 and RB10. RB6 interacts with both β-sheets through hydrogen bonds. The connection of the strands is such that the R domain is similar to the jelly-roll topology found in many proteins that are exclusively formed from antiparallel β-strands[30]. Jelly-roll domains include viral coat proteins, tumour necrosis factor, and the receptor-binding domain of ETA. The R domain differs somewhat from a strict jelly-roll topology (Fig. 4) in having strand 2 in the 'front' sheet, and having strand 10 in the 'back'. The R domain also is reminiscent of an immunoglobulin variable domain, but differs from the immunoglobulin fold in having an 'insert' of strands 5 and 6 between 4 and 7, and also in lacking two short strands (C′ and C″ in Fig. 4) between 4 and 5. The portion of the R domain that resembles a strict jelly-roll in topology is the right side as viewed in Fig. 4; and the portion that resembles the immunoglobulin variable domain is the left side, the side that is away from the rest of the DT monomer. Conceivably it is this immunoglobulin variable-like moiety that is involved in receptor recognition.

The DT dimer. Two monomers associate tightly to form a dimer with an interface between RB1/RB2 of one DT molecule and RB2/RB1 of the other DT molecule related by 2-fold rotational symmetry (Fig. 1*d*). This interface is one of three major protein–protein contacts in crystal packing and involves three hydrogen bonds per monomer. These hydrogen bonds are well defined because they are formed between main chain N and C atoms

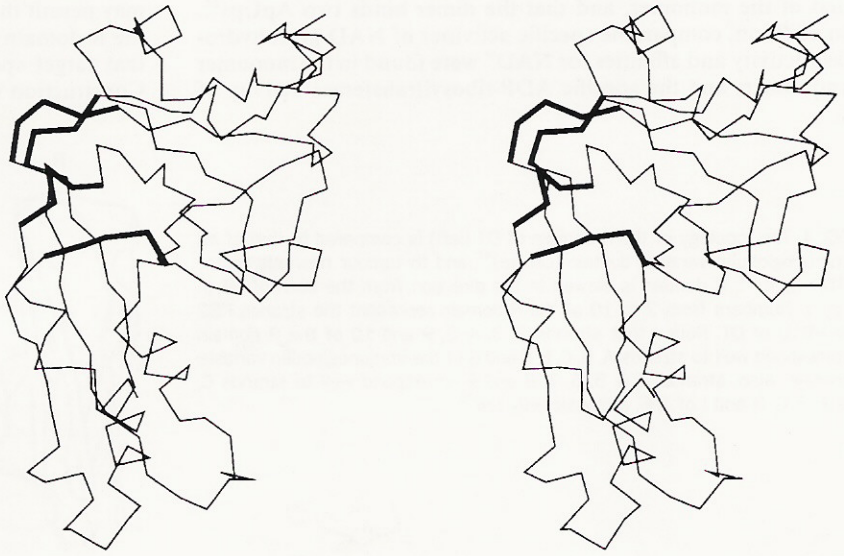


FIG. 2 Stereo pair of the Cα skeleton of the C domain. The entrance to the active site is at the lower right. The four loops, CL1 to CL4, are highlighted. Notice that they form a hinge which may permit the C domain to form a more elongated structure.

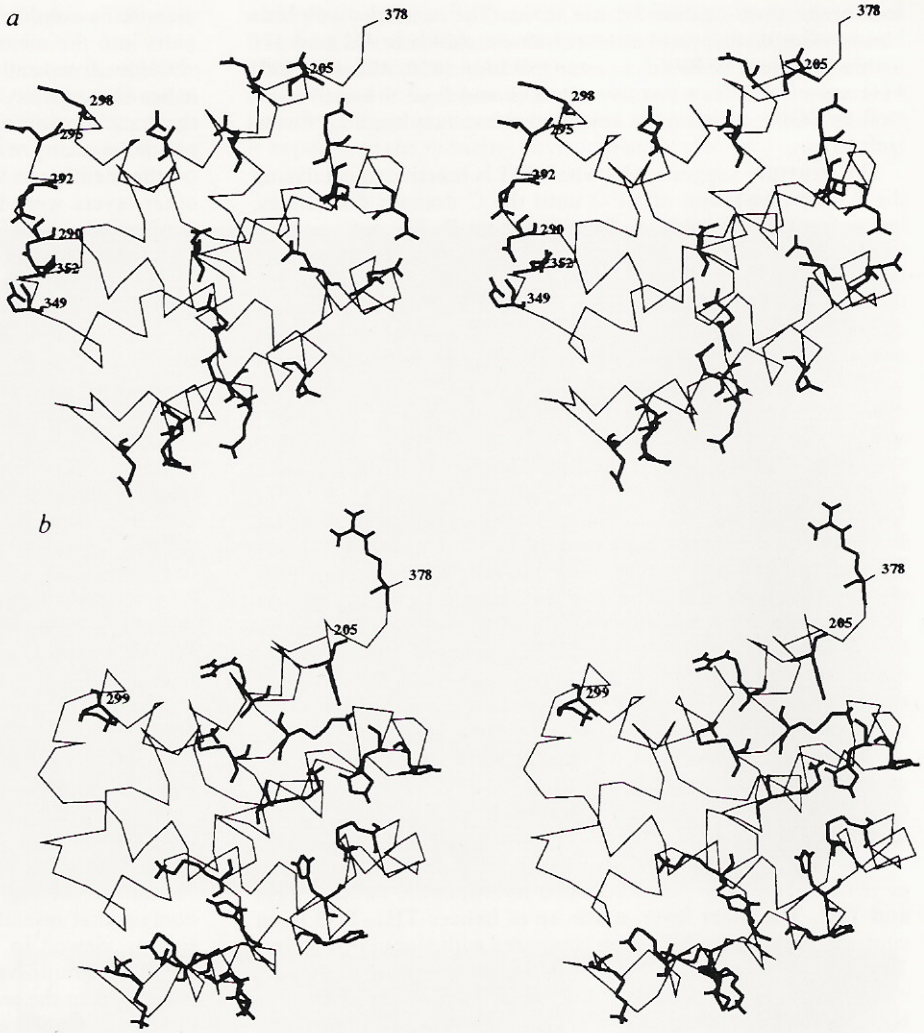


FIG. 3 Stereo pair of the Cα skeleton of the T domain, with the direction of view from the right side of DT in Fig. 1. Helix TH1 lies at the back, starting at residue 205. Helix TH2 runs to the left at the bottom, followed by a turn and helix TH3 running to the right. At front centre is TH5 (running to the left) and above it are helices TH6 and TH7. Behind these pairs of antiparallel helices is another pair of antiparallel helices, TH8 and TH9, with TH9 running upwards and ending at residue 378. *a*, Asp and Glu side chains are shown. Notice the tips of two helix layers, TL3 and TL5 contain a total of six acidic groups (on the left). *b*, Lys, Arg and His side chains are shown. Notice the positive charge asymmetry, with all charges at the bottom and back of the domain, except Lys 299 near loop TL3 between TH5 and TH6.

of RB1 and RB2. The other interfaces are not common among three different crystal forms. The inability of the dimer to bind to the DT receptor[16] suggests that the dimer interaction sterically blocks the receptor-binding domains of each monomer from the receptors on the surface of a target cell. The conformational differences between the monomer in the dimer and the native monomeric DT remain uncertain, but biochemical evidence suggests they are not large. Binding data show the affinity constant of the dimer for ApUp is the same as that of the monomer, and that the dimer binds two ApUps[16]. In addition, comparable specific activities of NAD-glycohydrolase activity and affinities for NAD^+ were found in the monomer and dimer; and the specific ADP-ribosyltransferase activity of fragment A released from the dimer after reduction was the same as that from the monomer[16]. These findings show that the conformation of the C domain, and of the portion of the R domain interfacing the C domain, are relatively unperturbed in the dimer.

Chimaeric toxins. The three-dimensional structure of DT demonstrates that the receptor-binding function is associated with a discrete, compact domain and defines the boundary between the R domain and the C and T domains. This finding may permit the design of re-engineered DT molecules in which the R domain is replaced by immunoglobulin or other domains that target specific cell types for killing by chimaeric toxins. Construction of a hybrid protein using gene splicing has been

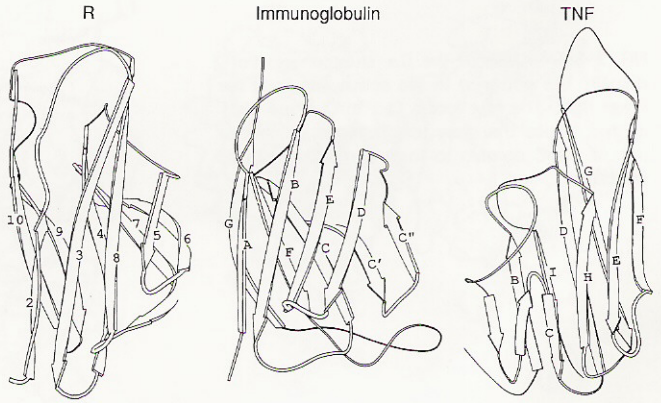


FIG. 4 The topology of the R domain of DT (left) is compared to that of an immunoglobulin variable domain (centre)[47] and to tumour necrosis factor (TNF, right)[48]. R domain is viewed in the direction from the back of DT in Fig. 1. Numbers from 2 to 10 of the R domain represent the strands RB2 to RB10 of DT. Notice that strands 2, 3, 4, 8, 9 and 10 of the R domain correspond well to strands A, B, C, E, F and G of the immunoglobulin variable domain. Also, strands 3, 4, 5, 6, 7, 8 and 9 correspond well to strands C, D, E, F, G, H and I of TNF, a classic jelly-roll[49].

reported[50], where the R domain was nearly replaced by an interleukin-2. Elucidation of the boundary between the C and T domains and the R domain may now facilitate the design of more effective chimaeras of DT. □

Received 17 February; accepted 30 March 1992.

1. Greenfield, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80,** 6853–6857 (1983).
2. Moskaug, J. O., Sletten, K., Sandvig, K. & Olsnes, S. *J. biol. Chem.* **264,** 15709–15713 (1989).
3. Collier, R. J. & Kandel, J. *J. biol. Chem.* **246,** 1496–1503 (1971).
4. Sandvig, K. & Olsnes, S. *J. biol. Chem.* **256,** 9068–9076 (1981).
5. Yamaizumi, M., Mekada, E., Uchida, T. & Okada, Y. *Cell* **15,** 245–250 (1978).
6. Colombatti, M., Dell'Arciprete, L., Rappuoli, R. & Tridente, G. *Meth. Enzym.* **178,** 404–422 (1989).
7. Williams, D. P. *et al. Prot. Engng* **1,** 493–498 (1987).
8. Chaudhary, V. K., Gallo, M. G., FitzGerald, D. J. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **87,** 9491–9494 (1990).
9. Greenfield, L., Johnson, V. G. & Youle, R. J. *Science* **238,** 536–539 (1987).
10. Papini, E. *et al. Eur. J. Biochem.* **169,** 637–644 (1987).
11. Moskaug, J. O., Stenmark, H. & Olsnes, S. *J. biol. Chem.* **266,** 2652–2659 (1991).
12. Blewitt, M. G., Chung, L. A. & London, E. *Biochemistry* **24,** 5458–5464 (1985).
13. Kagan, B. L., Finkelstein, A. & Colombini, M. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4950–4954 (1981).
14. Donovan, J. J., Simon, M. I., Draper, R. K. & Montal, M. *Proc. natn. Acad. Sci. U.S.A.* **78,** 172–176 (1981).
15. Collier, R. J., Westbrook, E. M., McKay, D. B. & Eisenberg, D. *J. biol. Chem.* **257,** 5283–5285 (1982).
16. Carroll, S. F., Barbieri, J. T. & Collier, R. J. *Biochemistry* **25,** 2425–2430 (1986).
17. Fujii, G., Choe, S., Bennett, M. J. & Eisenberg, D. *J. molec. Biol.* **222,** 861–864 (1991).
18. Wang, B.-C. *Meth. Enzym.* **115,** 90–112 (1985).
19. Brünger, A. T. *Acta crystallogr.* A**47,** 195–204 (1991).
20. Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **15,** 24–31. (1962)
21. Zhang, K. Y. J. & Main, P. *Acta crystallogr.* A**46,** 377–381 (1991).
22. Lüthy, R., Bowie, J. U. & Eisenberg, D. *Nature* **356,** 83–85 (1992).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* A**47,** 110–119 (1991).
24. Brünger, A. T. *Nature* **355,** 472–475 (1992).
25. Allured, V. S., Collier, R. J., Carroll, S. F. & McKay, D. B. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1320–1324 (1986).
26. Carroll, S. F. & Collier, R. J. *Molec. Microbiol.* **2,** 293–296 (1988).
27. Brandhuber, B. J., Allured, V. S., Falbel, T. G. & McKay, D. B. *Proteins* **3,** 146–154 (1988).
28. Sixma, T. K. *et al. Nature* **351,** 371–377 (1991).
29. Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. *Nature* **337,** 93–96 (1989).
30. Richardson, J. *Adv. Prot. Chem.* **34,** 167–339 (1981).
31. Li, J. D., Carroll, J. & Ellar, D. J. *Nature* **353,** 815–821 (1991).
32. Carroll, S. F. & Collier, R. J. *Proc. natn. Acad. Sci. U.S.A.* **81,** 3307–3311 (1984).
33. Papini, E., Schiavo, G., Sandona, D., Rappuoli, R. & Montecucco, C. *J. biol. Chem.* **264,** 12385–12388 (1989).
34. Papini, E. *et al. J. biol. Chem.* **266,** 2494–2498 (1991).
35. Giannini, G., Rappuoli, R. & Ratti, G. *Nucleic Acids Res.* **12,** 4063–4069 (1984).
36. Collins, C. M. & Collier, R. J. *Biochim. biophys. Acta* **828,** 138–143 (1985).
37. Zhao, J.-M. & London, E. *Biochemistry* **27,** 3398–3403 (1988).
38. Proia, R. L., Wray, S. K., Hart, D. A. & Eidels, L. *J. biol. Chem.* **255,** 12025–12033 (1980).
39. Lory, S. & Collier, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77,** 267–271 (1980).
40. Rees, D. C., DeAntonio, L. & Eisenberg, D. *Science* **245,** 510–513 (1989).
41. McLaughlin, S. *Curr. Top. Memb. Trans.* **9,** 71–144 (1977).
42. Weissman, L. thesis, Univ. California (1979).
43. Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. *Acta crystallogr.* A**43,** 1–5 (1987).
44. Jones, A. T. *Meth. Enzym.* **115,** 157–171 (1985).
45. Ponder, J. W. & Richards, F. M. *J. molec. Biol.* **193,** 775–791 (1987).
46. Brünger, A. T. & Krukowski, A. *Acta crystallogr.* A**46,** 585–593 (1990).
47. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **141,** 369–391 (1980).
48. Eck, M. J. & Sprang, S. R. *J. biol. Chem.* **264,** 17595–17605 (1989).
49. Jones, E. Y., Stuart, D. I. & Walker, N. P. C. *Nature* **338,** 225–228 (1989).
50. Kiyokawa, T., Williams, D. P., Snider, C. E., Strom, T. B. & Murphy, J. R. *Prot. Engng* **4,** 463–468 (1991).

ACKNOWLEDGEMENTS. We thank E. M. Westbrook, B. Dijkstra, S. F. Carroll and J. T. Barbieri for much work and D. B. McKay for providing the coordinates of ETA, the NIH for support, and a traineeship for M.J.B